



## Entropy and Average Cost of AUH Codes

Valeriu Munteanu, Daniela Tarniceriu, Gheorghe Zaharia

### ► To cite this version:

Valeriu Munteanu, Daniela Tarniceriu, Gheorghe Zaharia. Entropy and Average Cost of AUH Codes. Applied Mathematics & Information Sciences, 2013, 1 (1), pp.1-7. hal-00776585

**HAL Id: hal-00776585**

**<https://hal.science/hal-00776585>**

Submitted on 15 Jan 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Paper title: ENTROPY AND AVERAGE COST OF AUH CODES**

### **Authors:**

#### **1. Valeriu Munteanu**

Academic title: Professor Ph.D. Engineer

*Affiliation: Technical University “Gh. Asachi” Iasi, Faculty of Electronics, Telecommunications and Information Technology*

*Correspondence address: Bd. Carol I no. 11, 700506, Iasi, Romania.*

Phone: (+40) – 232 – 213737 ext 228

Fax: (+40) – 0232 – 217720

email: [vmuntean@etti.tuiasi.ro](mailto:vmuntean@etti.tuiasi.ro)

#### **2. Daniela Tarniceriu**

Academic title: Professor Ph.D. Engineer

*Affiliation: Technical University “Gh. Asachi” Iasi, Faculty of Electronics, Telecommunications and Information Technology*

*Correspondence address: Bd. Carol I no. 11, 700506, Iasi, Romania.*

Phone: (+40) – 232 – 213737 ext 235

Fax: (+40) – 0232 – 217720

email: [tarniced@etti.tuiasi.ro](mailto:tarniced@etti.tuiasi.ro)

#### **3. Gheorghe Zaharia**

Academic titles: Assistant Professor Ph.D. Engineer

*Affiliation: Universite Europeenne de Bretagne, INSA – Rennes, France.*

*Correspondence address: 20, Av. des Buttes de Coesmes, Rennes, FR 35043 France*

Phone : (+33)223238589

Fax : (+33)223238439

email: [gheorghe.zaharia@insa-rennes.fr](mailto:gheorghe.zaharia@insa-rennes.fr)

# ENTROPY AND AVERAGE COST OF AUH CODES

Valeriu Munteanu<sup>1</sup>, Daniela Tarniceriu<sup>1</sup>, Gheorghe Zaharia<sup>2</sup>

<sup>1</sup>“Gheorghe Asachi” Technical University, Faculty of Electronics, Telecommunications and Information

Technology, Department of Telecommunications, Bd. Carol I, 700506, Iasi, Romania

<sup>2</sup> IETR – INSA, UMR CNRS 6164 Rennes, France,

emails: [vmuntean@etti.tuiasi.ro](mailto:vmuntean@etti.tuiasi.ro), [tarniced@etti.tuiasi.ro](mailto:tarniced@etti.tuiasi.ro), [Gheorghe.Zaharia@insa-rennes.fr](mailto:Gheorghe.Zaharia@insa-rennes.fr)

*Abstract*— In this paper we address the class of anti-uniform Huffman (AUH) codes, named also unary codes, for sources with finite and infinite alphabet, respectively. Geometric, quasi-geometric, Fibonacci, exponential, Poisson, and negative binomial distributions lead to anti – uniform sources for some ranges of their parameters. Huffman coding of these sources results in AUH codes. We prove that as result of this encoding, in general, sources with memory are obtained. For these sources we attach the graph and determine the transition matrix between states, the state probabilities and the entropy. If  $c_0$  and  $c_1$  denote the costs for storing or transmission of symbols “0” and “1”, respectively, we compute the average cost for these AUH codes.

Keywords: Huffman coding, average codeword length, average cost, entropy.

## 1. Introduction

Let  $(p_1, p_2, \dots, p_n)$  be the probability distribution of a  $n$  messages source  $\xi_n = \{s_1, s_2, \dots, s_n\}$ . It is well known that the Huffman encoding algorithm [1] produces an optimal binary prefix-free code for  $\xi_n$ . A binary Huffman code is usually represented by a binary tree, whose leaves correspond to the source messages. The two edges emanating from each intermediate tree node (father) are labeled either 0 or 1. For related literature on Huffman coding and Huffman

trees, we refer the reader to [2]–[6]. The length between the root and a leaf is the length of the binary codeword associated with the corresponding message.

Assuming that  $v_i, i = 1, 2, \dots, n$ , is the codeword representing the message  $s_i$ , we denote the length of  $v_i$  by  $l_i$ . The optimality of Huffman coding implies that  $l_i \leq l_j$ , if  $p_i > p_j$ .

Anti uniform Huffman (AUH) codes were firstly introduced in [7]. A Huffman code representing a finite source  $\xi_n$  satisfying  $p_1 \geq p_2 \geq \dots \geq p_n > 0$  is an anti-uniform code, if  $l_i = i, i = 1, 2, \dots, n-1$  and  $l_n = l_{n-1}$ . A source  $\xi_n$  having an anti-uniform Huffman code is called an anti-uniform source. These sources were extensively analysed, concerning bounds on average codeword length, entropy and redundancy for different types of probability distribution [7]–[9]. The AUH sources appear in a wide variety of situations in the real world, because this class of sources have the property of achieving minimum redundancy in different situations and minimal average cost in highly unbalanced cost regime [10]–[12]. These properties determine a wide range of applications and motivate us to study these sources from an information theoretic perspective. One example is the telegraph channel with the alphabet  $\{., -\}$  in which dashes are twice as long as dots [13]. Another is the  $\{a, b\}$  run – length – limited codes used in magnetic and optical storage, in which the binary codewords are constrained so that each 1 must be preceded by at least  $a$ , and at most  $b$ , 0's [14]. The binary Huffman codes, constrained so all codewords must end in a 1, are used for group testing and self-synchronizing codes [15, 16]. As another example, binary codes whose codewords contain at most a specified number of 1's are used for energy minimization of transmissions in mobile environments [17]. AUH sources can be generated by several probability distributions. It has been shown that geometric, quasi-geometric, Fibonacci, exponential, Poisson and negative binomial distributions lie in the class of AUH sources for some regimes of their parameters [7], [18], [19], [20]. Related topic was addressed in [21], where the authors studied weakly super increasing (WSI) and partial WSI sources in connection with Fibonacci numbers and golden mean, which appeared extensively in modern science and, in particular, have applications in coding and information theory.

The rest of the paper is organized as follows. In Section 2 we present the Huffman encoding of an anti-uniform source and the graph of the source with memory resulting by binary Huffman encoding of AUH sources. We show that, in general, employing Huffman coding, a source with memory results. The entropy and the average cost of the code are also derived. In Sections 3 we compute the code entropy, as well as the average cost for AUH codes corresponding to sources with geometric, quasi-geometric, Fibonacci, exponential, Poisson and negative binomial distributions, respectively. Finally, we conclude the paper in Section 4.

## 2. Entropy and average cost of AUH codes

Let us consider a discrete and memoryless source, characterized by the distribution:

$$\xi_n : \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}, \quad (1)$$

$$p_1 \geq p_2 \geq \dots \geq p_n \quad (2)$$

$$\sum_{i=1}^n p_i = 1 \quad (3)$$

If [7]

$$\sum_{k \geq i+2}^n p_k \leq p_i, 1 \leq i \leq n-3, \quad (4)$$

the source becomes anti-uniform.

After a binary Huffman encoding of the source with the distribution in (1) that fulfils (4), the graph in Fig. 1 is obtained.

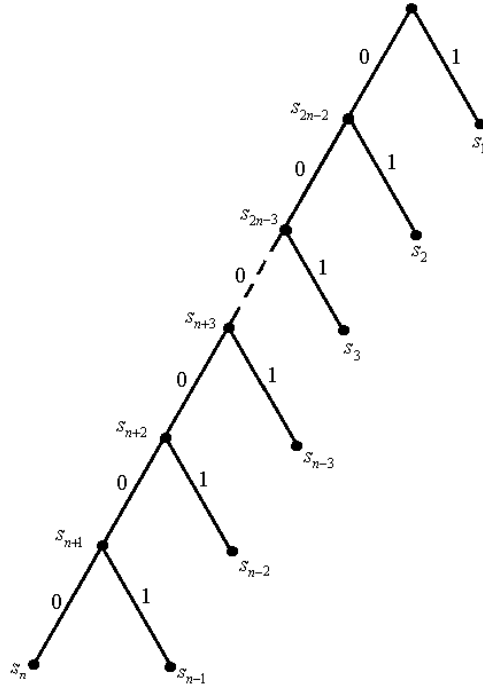


Fig. 1. The graph of binary Huffman encoding for the source  $\xi_n$  with distribution in (1)

The structure of codewords resulting from binary Huffman encoding is:

$$\begin{aligned}
 s_0 &\rightarrow v_0 \rightarrow 1 \\
 s_1 &\rightarrow v_1 \rightarrow 01 \\
 s_2 &\rightarrow v_2 \rightarrow 001 \\
 &\dots\dots\dots \\
 s_{n-1} &\rightarrow v_{n-1} \rightarrow \underbrace{00\dots01}_{n-2} \\
 s_n &\rightarrow v_n \rightarrow \underbrace{00\dots00}_{n-1}
 \end{aligned}$$

The length  $l_i$  of the codeword associated with the message  $s_i$  is the number of edges on the path between the root and the node  $s_i$  in the Huffman tree.

$$l_i = i, \quad i = 1, 2, \dots, n-1 \quad (5)$$

$$l_n = l_{n-1} \quad (6)$$

The average codeword length is determined with

$$\bar{l}_n = \sum_{i=1}^n p_i l_i = \sum_{i=1}^{n-1} i p_i + (n-1) p_n. \quad (7)$$

In Fig. 1  $s_{n+i}, i = 1, 2, \dots, n-2$  denote the intermediate nodes in the graph, also called parents.

The probability of a parent is obtained as the sum of the two sibling probabilities. Denoting by  $p_{n+i}$  the probabilities of intermediate nodes, we have

$$p_{n+i} = \sum_{k=n-i}^n p_k; i = 1, 2, \dots, n-2 \quad (8)$$

For a sequence of messages  $s_i$  the source delivers a string of binary symbols from the code alphabet  $X = \{x_0 = 0, x_1 = 1\}$ . As the probabilities of symbols in the binary string depend on the node from which they are generated, the set  $X$ , which is the output bitstream obtained as result of binary Huffman coding is a source with memory.

When a terminal node  $s_i, i = 1, 2, \dots, n$  is reached, the source  $\xi_n$  will deliver another message and the source with memory  $X$  will generate another string of binary symbols.

The graph attached to the source with memory  $X$  can be obtained from the Huffman encoding graph in Fig. 1, as follows:

- a) We link the terminal nodes with the graph root;
- b) The branches between successive nodes have the probabilities equal to the ratio between the probability of the node in which the branch ends and the probability of the node from which it starts, excepting the branches linking the terminal nodes with graph root, whose probabilities are equal to unity;
- c) Each terminal node  $s_i, i = 1, 2, \dots, n$ , or intermediate ones  $s_{n+i}, i = 1, 2, \dots, n-2$ , (excepting the root of the encoding graph) will represent the states  $S_i, i = 1, 2, \dots, 2n-2$ , of the source with memory  $X$ . The set of these states is denoted by  $S = \{S_1, S_2, \dots, S_{2n-2}\}$ . The graph of the source  $X$  is shown in Fig. 2.

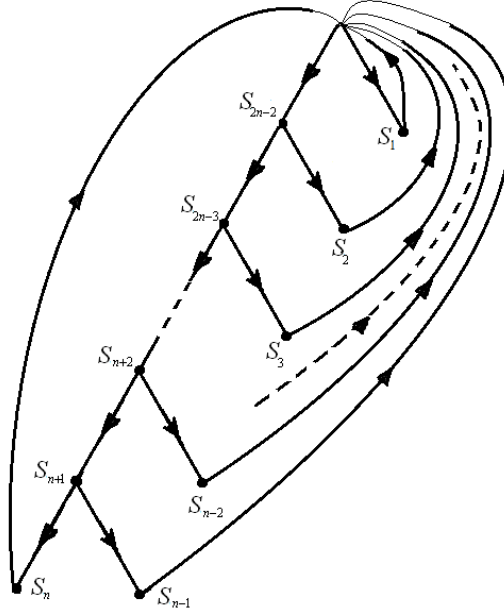


Fig. 2 The graph of the source with memory X

The transition probabilities from the state  $S_i$  in the state  $S_j$ , that is,  $p(S_j | S_i)$ , is equal to the probability of the branch between the node  $S_i$  and the node  $S_j$ . When the source enters the state  $S_j$  from the state  $S_i$ , it generates either the symbol  $x_0 = 0$ , or  $x_1 = 1$ . Therefore, the probability of delivering the symbol  $x_j, j = 0, 1$ , from the state  $S_i$  is the same as the probability to reach the state  $S_j$ , starting from the state  $S_i$ , that is:

$$p(x_1 | S_i) = p(S_1 | S_i) = p_1, i = 1, 2, \dots, n \quad (9)$$

$$p(x_0 | S_i) = p(S_{2n-2} | S_i) = \sum_{k=2}^n p_k, i = 1, 2, \dots, n \quad (10)$$

$$p(x_1 | S_{n+i}) = p(S_{n-i} | S_{n+i}) = \frac{p_{n-i}}{\sum_{k=n-i}^n p_k}, i = 1, 2, \dots, n-2 \quad (11)$$

$$p(x_0 | S_{n+i}) = p(S_{n+i-1} | S_{n+i}) = \frac{\sum_{k=n-i+1}^n p_k}{\sum_{k=n-i}^n p_k}, i = 1, 2, \dots, n-2 \quad (12)$$

Considering (9), (10), (11) and (12) as well as the graph in Fig. 2, the transition matrix between states is:



$$\mathbf{T} = \begin{array}{c} \begin{array}{cccccccccccc} S_1 & S_2 & \cdots & S_{n-2} & S_{n-1} & S_n & S_{n+1} & S_{n+2} & \cdots & S_{2n-3} & S_{2n-2} \end{array} \\ \left[ \begin{array}{cccccccccccc} p_1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \sum_{k=2}^n p_k \\ p_1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \sum_{k=2}^n p_k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \sum_{k=2}^n p_k \\ 0 & 0 & \cdots & 0 & \frac{p_{n-1}}{\sum_{k=n-1}^n p_k} & \frac{\sum_{k=n}^n p_k}{\sum_{k=n-1}^n p_k} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \frac{p_{n-2}}{\sum_{k=n-2}^n p_k} & 0 & 0 & \frac{\sum_{k=n-1}^n p_k}{\sum_{k=n-2}^n p_k} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \frac{p_2}{\sum_{k=2}^n p_k} & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{\sum_{k=3}^n p_k}{\sum_{k=2}^n p_k} & 0 \end{array} \right] \begin{array}{c} S_1 \\ S_2 \\ \vdots \\ S_n \\ S_{n+1} \\ S_{n+2} \\ \vdots \\ S_{2n-2} \end{array} \end{array} \quad (13)$$

where the entries of the matrix  $\mathbf{T}$  are  $t_{ij}=p(S_j|S_i)$ .

Let  $\pi_i, i=1,2,\dots,2n-2$ , denote the stationary state probabilities of the source with memory. They can be determined by means of [22]

$$[\pi_1, \pi_2, \dots, \pi_{2n-2}] = [\pi_1, \pi_2, \dots, \pi_{2n-2}] \mathbf{T} \quad (14)$$

$$\sum_{i=1}^{2n-2} \pi_i = 1 \quad (15)$$

Considering (7) and (13), from (14) and (15) we obtain the state probabilities as:

$$\pi_i = \frac{p_i}{l_n}, i=1,2,\dots,n \quad (16)$$

$$\pi_{n+i} = \frac{1}{l_n} \sum_{k=n-i}^n p_k, i=1,2,\dots,n-2 \quad (17)$$

Generally, the entropy of the source with memory is computed by [23]

$$H(X) = - \sum_{i=1}^{2n-2} \sum_{j=1}^2 \pi_i p(x_j | S_i) \log p(x_j | S_i) \quad (18)$$

where  $\pi_i$  are given in (16) and (17) and  $p(x_j | S_i)$  are given in (9), (10), (11) and (12).

Let  $c_0$  and  $c_1$  be the costs of storing or transmission of symbols “0” and “1”, respectively, resulted after the binary Huffman encoding of source  $\xi_n$ . The average cost is determined by [9]

$$\bar{C} = \sum_{i=1}^n p_i [n_0(i)c_0 + n_1(i)c_1], \quad (19)$$

where  $n_0(i)$  and  $n_1(i)$  denote the number of 0's and 1's in the codeword  $c_i$ .

Considering the structure of codewords for AUH sources, (5) and (6), the average cost is

$$\bar{C} = \sum_{i=1}^n p_i [(i-1)c_0 + c_1] + (n-1)c_0 p_n \quad (20)$$

### 3. Case studies for distributions leading to AUH sources

For all distributions we considered, we are interested to found out the condition imposed to the source parameter so that it is AUH, to derive the code entropy and the average cost.

#### 3.1. AUH sources with geometric distribution

Let there be a discrete source characterized by the geometric distribution:

$$\xi_n : \begin{pmatrix} s_1 & s_2 & s_3 & \cdots & s_n \\ p_1 = 1-p & p_2 = p(1-p) & p_3 = p^2(1-p) & \cdots & p_n = p^{n-1}(1-p) \end{pmatrix}, \quad (21)$$

In this case the source is not complete, because

$$\sum_{i=1}^n p_i = 1 - p^n. \quad (22)$$

To form a complete source, we normalize the probabilities  $p_i$  in (21), resulting the complete source with distribution:

$$\xi_n^{norm} : \begin{pmatrix} s_1 & s_2 & s_3 & \cdots & s_n \\ p_1 = \frac{1-p}{1-p^n} & p_2 = \frac{p(1-p)}{1-p^n} & p_3 = \frac{p^2(1-p)}{1-p^n} & \cdots & p_n = \frac{p^{n-1}(1-p)}{1-p^n} \end{pmatrix} \quad (23)$$

For this source to become AUH, relation (4) is required to be fulfilled. Replacing the probabilities from (23) in (4), we have [18]:

$$0 < p \leq \frac{\sqrt{5}-1}{2} \quad (24)$$

The average codeword length results by replacing the probabilities  $p_i$  from (23) into (7).

$$\overline{l}_n^{norm} = \frac{1}{(1-p)(1-p^n)} \left[ 1 - p^{n-1} - (n-1)p^n + (n-1)p^{n+1} \right] \quad (25)$$

### Theorem

The entropy and the average cost of the source with memory resulted by binary encoding of the AUH source with Poisson distribution are determined by:

$$H_n(X) = -\frac{1}{\overline{l}_n^{norm}} \left[ \log(1-p^n) - \log(1-p) - \frac{1-np^{n-1}+(n-1)p^n}{(1-p^n)(1-p)} p \log p \right] \quad (26)$$

$$\overline{C}_n = \frac{1}{1-p^n} \left[ (1-p^{n-1})c_0 + \frac{1-np^{n-1}+(n-1)p^n}{1-p} pc_1 \right] \quad (27)$$

### Proof

The stationary state probabilities are obtained by replacing the probabilities  $p_i$  from (23) into (16) and (17), and considering (25):

$$\left. \begin{aligned} \pi_i &= \frac{p^{i-1}(1-p)}{(1-p^n)\overline{l}_n^{norm}}, i=1,2,\dots,n \\ \pi_{n+i} &= \frac{p^{n-i-1}(1-p^{i+1})}{(1-p^n)\overline{l}_n^{norm}}, i=1,2,\dots,n-2 \end{aligned} \right\} \quad (28)$$

The probabilities to deliver the symbol  $x_0 = 0$  and  $x_1 = 1$ , from the states  $S_i, i=1,2,\dots,2n-2$ , result by replacing the probabilities  $p_i$  from (23) into (9) – (12):

$$\left. \begin{aligned} p(x_1 | S_i) &= \frac{1-p}{1-p^n}, i=1,2,\dots,n \\ p(x_0 | S_i) &= \frac{p(1-p^{n-1})}{1-p^n}, i=1,2,\dots,n \\ p(x_1 | S_{n+i}) &= \frac{1-p}{1-p^{i+1}}, i=1,2,\dots,n-2 \\ p(x_0 | S_{n+i}) &= \frac{p(1-p^i)}{1-p^{i+1}}, i=1,2,\dots,n-2 \end{aligned} \right\} \quad (29)$$

Substituting (28) and (29) into (18), the relation (26) results.

Substituting (23) into (20), the relation (27) results.

When the number of messages,  $n$ , of the source  $\xi_n$  increases, at limit, when  $n \rightarrow \infty$ , we have

$$\bar{l}_\infty = \frac{1}{1-p} \quad (30)$$

$$H_\infty(X) = -p \log p - (1-p) \log(1-p) \quad (31)$$

$$\bar{C}_\infty = c_0 + \frac{p}{1-p} c_1 \quad (32)$$

### 3.2. AUH sources with quasi-geometric distribution

Let there be a discrete source characterized by the quasi-geometric distribution:

$$\xi_n : \begin{pmatrix} s_1 & s_2 & s_3 & \cdots & s_{n-1} & s_n \\ p_1 = 1-p & p_2 = \frac{p}{2} & p_3 = \frac{p}{2^2} & \cdots & p_{n-1} = \frac{p}{2^{n-2}} & p_n = \frac{p}{2^{n-1}} \end{pmatrix}, \quad (33)$$

Following similar procedures as in the previous case, we get:

- The range for the source parameter for the source to be AUH

$$0 < p \leq \frac{2}{3}. \quad (34)$$

- The average codeword length

$$\bar{l}_n = 1 + \frac{2^{n-2} - 1}{2^{n-3}} p \quad (35)$$

- The entropy of the source with memory  $X$

$$H_n(X) = -\frac{(1-p)^{n+1} \log(1-p) + [1 - np^{n-1} + (n-1)p^n] p \log p}{1 - p^{n-1} - (n-1)p^n + (n-1)p^{n+1}} \quad (36)$$

- The average cost of the source with memory  $X$

$$\bar{C}_n = \frac{(2^{n-1} - 1)pc_1 + (2^{n-2} - p)c_0}{2^{n-2}} \quad (37)$$

When  $n \rightarrow \infty$ , we have

$$H_\infty(X) = -\frac{1}{1+2p} [(1-p) \log(1-p) + p \log p - 2p] \quad (38)$$

$$\overline{C}_\infty = 2pc_1 + c_0 \quad (39)$$

In the special case, when  $p = \frac{1}{2}$ , the source  $\xi_n$  becomes dyadic one. In this case  $X$  becomes memoryless and then:

$$p(x_j | S_i) = \frac{1}{2}, j = 0, 1; i = 1, 2, \dots, 2n - 2, \quad (40)$$

$$\bar{l}_{dn} = 2 - \frac{1}{2^{n-2}} \quad (41)$$

$$H_{dn}(X) = 1 \quad (42)$$

$$\overline{C}_{dn} = \frac{(2^{n-1} - 1)(C_1 + C_0)}{2^{n-1}} \quad (43)$$

Imposing  $n \rightarrow \infty$  in (41) and (43), we have

$$\bar{l}_{d\infty} = 2 \quad (44)$$

$$\overline{C}_{d\infty} = c_0 + c_1 \quad (45)$$

### 3.3. AUH sources with Fibonacci distribution

Let there be a discrete and finite AUH source characterized by the Fibonacci distribution

$$\xi_n : \begin{pmatrix} s_1 & s_2 & \cdots & s_{n-2} & s_{n-1} & s_n \\ p_1 = \frac{f_{n-1}}{f_{n+1}} & p_2 = \frac{f_{n-2}}{f_{n+1}} & \cdots & p_{n-2} = \frac{f_2}{f_{n+1}} & p_{n-1} = \frac{f_1}{f_{n+1}} & p_n = \frac{f_1}{f_{n+1}} \end{pmatrix} \quad (46)$$

where  $f_n$  is the  $n^{th}$  Fibonacci number defined as

$$\left. \begin{aligned} f_1 &= f_2 = 1 \\ f_n &= f_{n-1} + f_{n-2}, n \geq 3 \end{aligned} \right\} \quad (47)$$

The source  $\xi_n$  is also AUH, because relation (4) is fulfilled.

- The average codeword length

$$\bar{l}_n = \frac{f_{n+3} - 3}{f_{n+1}} \quad (48)$$

- The entropy of the source with memory  $X$

$$H_n(X) = \frac{f_{n+1}}{f_{n+3}-3} \log f_{n+1} - \frac{1}{f_{n+3}-3} \sum_{i=1}^{n-1} f_i \log f_i \quad (49)$$

- The average cost of the source with memory  $X$

$$\bar{C}_n = \frac{f_{n+1}-f_1}{f_{n+1}} c_0 + \frac{f_{n+2}}{f_{n+1}} c_1 \quad (50)$$

### 3.4. AUH sources with exponential distribution

The density probability function of exponential distribution is

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (51)$$

We define the cumulative density function

$$F(i) = 1 - e^{-\lambda i}, \quad (52)$$

and the probabilities

$$p_i = F(i) - F(i-1) = e^{-(i-1)\lambda} (1 - e^{-\lambda}). \quad (53)$$

Let us consider the source  $\xi_n$  with exponential distribution

$$\xi_n : \begin{pmatrix} s_1 & s_2 & \cdots & s_{n-1} & s_n \\ p_1 = 1 - e^{-\lambda} & p_2 = e^{-\lambda}(1 - e^{-\lambda}) & \cdots & p_{n-1} = e^{-(n-2)\lambda}(1 - e^{-\lambda}) & p_n = e^{-(n-1)\lambda}(1 - e^{-\lambda}) \end{pmatrix} \quad (54)$$

- The range for the source parameter for the source to be AUH [8]

$$\lambda \geq \ln \left( \frac{\sqrt{5} + 1}{2} \right) \quad (55)$$

- The average codeword length

$$\bar{l}_n^{norm} = \frac{1}{(1 - e^{-\lambda})(1 - e^{-n\lambda})} \left[ 1 - e^{-(n-1)\lambda} - (n-1)e^{-n\lambda} + (n-1)e^{-(n+1)\lambda} \right] \quad (56)$$

- The entropy of the source with memory  $X$

$$H_n(X) = -\frac{1}{\bar{l}_n^{norm}} \left[ \log(1 - e^{-n\lambda}) - \log(1 - e^{-\lambda}) + \frac{1 - ne^{-(n-1)\lambda} + (n-1)e^{-n\lambda}}{(1 - e^{-n\lambda})(1 - e^{-\lambda})} \lambda e^{-\lambda} \log e \right] \quad (57)$$

- The average cost of the source with memory  $X$

$$\bar{C}_n = \frac{1}{1 - e^{-n\lambda}} \left[ (1 - e^{-(n-1)\lambda}) C_0 + \frac{1 - ne^{-(n-1)\lambda} + (n-1)e^{-n\lambda}}{1 - e^{-\lambda}} e^{-\lambda} C_1 \right] \quad (58)$$

When  $n \rightarrow \infty$ , we have

$$H_\infty(X) = -(1 - e^{-\lambda}) \left( \log(1 - e^{-\lambda}) + \frac{e^{-\lambda} \log e^{-\lambda}}{1 - e^{-\lambda}} \right) \quad (59)$$

$$\bar{C}_\infty = c_0 \frac{e^{-\lambda}}{1 - e^{-\lambda}} + c_1 \quad (60)$$

$$\bar{l}_\infty = \sum_{k=0}^{\infty} (k+1) e^{-k\lambda} (1 - e^{-\lambda}) = \frac{1}{1 - e^{-\lambda}}. \quad (61)$$

### 3.5 AUH sources with Poisson distribution

Let there be a discrete source with infinite alphabet, characterized by Poisson distribution:

$$\xi_n : \begin{pmatrix} s_0 & s_1 & s_2 & \cdots & s_n \\ p_0 = e^{-\lambda} & p_1 = e^{-\lambda} \frac{\lambda}{1!} & p_2 = e^{-\lambda} \frac{\lambda^2}{2!} & \cdots & p_n = e^{-\lambda} \frac{\lambda^n}{n!} \end{pmatrix}, \quad (62)$$

- The range of the parameter for the source to be AUH [18]

$$\lambda \leq 1 \quad (63)$$

- The average codeword length

$$\bar{l}_\infty = \lambda + 1. \quad (64)$$

The entropy of the source with memory  $X$

$$H_\infty(X) = \frac{1}{\lambda + 1} \left( \lambda \log e - \lambda \log \lambda + e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \log(n!) \right) \quad (65)$$

The average cost of the source with memory  $X$

$$\bar{C}_\infty = c_0 \lambda + c_1 \quad (66)$$

### 3.6. AUH sources with negative binomial distribution

Let there be a discrete source with infinite alphabet, characterized by the negative binomial distribution:

$$\xi_n : \begin{pmatrix} s_0 & s_1 & s_2 & \cdots & s_n \\ p_0 = C_{r-1}^{r-1} p^r & p_1 = C_r^{r-1} p^r q & p_2 = C_{r+1}^{r-1} p^r q^2 & \cdots & p_n = C_{r+n-1}^{r-1} p^r q^n \end{pmatrix}, \quad (67)$$

where  $q=1-p$ .

- The average codeword length

$$\bar{l}_\infty = \sum_{n=0}^{\infty} (n+1) C_{r+n-1}^{r-1} p^r q^n = \frac{rq+p}{p}. \quad (68)$$

The entropy of the source with memory  $X$

$$H_\infty(X) = -\frac{p}{rq+p} \left( r \log p + \frac{qr}{p} \log q + p^r \sum_{n=1}^{\infty} C_{r+n-1}^{r-1} q^n \log C_{r+n-1}^{r-1} \right) \quad (69)$$

The average cost of the source with memory  $X$

$$\bar{C}_\infty = c_0 \frac{qr}{p} + c_1 \quad (70)$$

## 7. Conclusions

In this paper we have considered the class of AUH sources with finite and infinite alphabets. Performing a binary Huffman encoding of these sources, we show that sources with memory result. For these sources we build the encoding graph and the graph of the source with memory  $X$ . The graph of the source with memory is obtained from the encoding graph by linking the terminal nodes with the graph root. The states of the source with memory correspond to the terminal or intermediate nodes in the encoding graph, excepting the root. We determined in the general case the state probabilities of the source with memory, as well as the transition probabilities between states. The entropy of this source with memory is computed. We assumed the costs  $c_0$  and  $c_1$  for the symbols “0” and “1”, respectively, and compute the average cost for these Huffman codes. Obviously, the Huffman encoding procedure assures minimum average length, but the average cost is not minimum. It can be easily verified that if the costs of symbols “0” and “1” are equal to unity,



the average cost becomes equal to the average length. We applied the results for several AUH sources with geometric, quasi-geometric, Fibonacci, exponential, Poisson and negative binomial distributions. We have also analyzed the limit cases, when the source alphabet increases unlimited, that is,  $n \rightarrow \infty$ .

## References

- [1] R. Huffman, "A method for the construction of minimum – redundancy codes," *Proc. IRE*, vol. 40, pp. 1098 – 1101, 1952.
- [2] T. Linder, V. Tarokh, K. Seeger, "Existence of optimal prefix codes for infinite source alphabets," *IEEE Trans. Inf. Theory*; vol. 43, No. 6, pp. 2026-2028, Nov. 1997
- [3] R. M. Capocelli, A. De Santis, "A note on D-ary Huffman codes," *IEEE Trans. Inf. Theory* vol. 37, No.1, pp. 174-191, Jan. 1991.
- [4] R. Gallager, "Variations by a theme of Huffman," *IEEE Trans. Inf. Theory* vol. 24, No. 6, pp. 668-74, Nov. 1978.
- [5] O. Johnsen, "On the redundancy of binary Huffman codes," *IEEE Trans. Inf. Theory*, vol. 26, No. 2, pp. 220-222, March, 1980.
- [6] M. Khorsavifard, M. Esmaeili, H. Saidi, T.A. Gulliver, "A tree based algorithm for generating all possible binary compact codes with N codewords," *IEICE Trans. Fundam. Electron. Commun. Computer Sci.*, vol. E86-A, No. 10, pp. 2510-2516, Oct. 2003.
- [7] M. Esmaeili, A. Kakhbod, "On antiuniform and partially antiuniform sources," *Proc. IEEE ICC*, pp. 1611 – 1615 June 2006
- [8] S. Mohajer, A. Kakhbod, "Anti – uniform Huffman codes," *IET Commun.*, vol.5, pp. 1213 – 1219 2011.
- [9] S. Mohajer, A Kakhbod, "Tight bounds on the AUH codes," Information Sciences and Sysyems, 2008. CISS 2008. 42<sup>nd</sup> Annual Conference on, pp. 1010 – 1014, 19 – 21 March
- [10] S. Mohajer, P. Pakzad, A. Kakhbod, "Tight bounds on the redundancy of Huffman codes," *Proc. IEEE ITW*, , pp. 131 – 135, March, 2006.
- [11] O. Johnsen, "On the redundancy of binary Huffman codes," *IEEE Trans. Inf. Theory*. vol. 26, pp.220-222, 1980.
- [12] P. Bradford, M Golin, L. L. Larmore and W. Rytter, "Optimal prefix free codes for unequal letter costs and dynamic programming with the Monge property," *J. Algorithms*, vol. 42, pp. 277 – 303, 2002.
- [13] E. N. Gilbert, "Coding with digits of unequal costs," *Trans. Inf. Theory*. vol. 41, pp. 596-600, 1995.
- [14] M. Golin, G. Rote, "A dynamic programming algorithm for constructing optimal prefix-free codes for unequal costs," *IEEE Trans. IT* 44(5):1770-1781, 1998.
- [15] T. Berger, R. Yeung, "Optimum 1-ended binary prefix codes," *Trans. Inf. Theory*. vol.36 No.6, pp. 1434-1441, Nov. 1990.
- [16] A. De Santis, R. M. Capocelli, G. Persiano "Binary prefix codes ending in a 1 *Trans. Inf. Theory*. vol. 40, pp. 1296-1302, July 1994.
- [17] E. Korach, S. Dolov and D. Yukelson, "The sound of silence: Guessing games for saving energy in mobile environment," In *Proc. Of the Eighteenth Annual Joint Conference of IEEE Computer and Communications Societies (IEEE INFOCOM'99)*, pp. 768-775, 1999.
- [18] M. Esmaeili, A.Kakhbod, "On information theory parameters of infinite anti-uniform sources," *IET Commun.*, vol.1, pp. 1039 – 1041, 2007.
- [19] P. Humblet, "Optimal source coding for a class of integer alphabets," *IEEE Trans. Inf. Theory*, vol 24, No.1, pp. 110 – 112, 1978.
- [20] R. Gallager, D. Van Voorhis, "Optimal source coding for geometrically distributed integer alphabets," *IEEE Trans. Inf. Theory*, vol. 21, No. 2, pp. 228 – 230, 1975.
- [21] M. Esmaeili, A. Gulliver and A. Kakhbod, "The Golden Mean, Fibonacci Matrices and Partial Weakly Super-Increasing Sources," *Chaos, Solitions and Fractals*, Vol. 42, No. 1, pp. 435 – 440, 2009.
- [22] J. G. Kemeny, T. L. Snell, Finite Markov Chains, *Princeton, NJ: Van Nostrand*, 1960.
- [23] T. M. Cover, J. A. Thomas, Elements of Information Theory. *John Wiley and Sons, Inc.* New York, 1991.